

STRUCTURED SPARSITY THROUGH REWEIGHTING AND APPLICATION TO DIFFUSION MRI

Anna Auría^{*}, A. Daducci^{*,†}, J.P. Thiran^{*,†}, Y. Wiaux[‡]

^{*}Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

[†] University of Lausanne (UNIL), CH-1011 Lausanne, Switzerland

[‡] Institute of Sensors, Signals, and Systems, Heriot-Watt University, EH14 4AS, UK

ABSTRACT

We consider the problem of multiple correlated sparse signals reconstruction and propose a new implementation of structured sparsity through a reweighting scheme. We present a particular application for diffusion Magnetic Resonance Imaging data and show how this procedure can be used for fibre orientation reconstruction in the white matter of the brain. In that framework, our structured sparsity prior can be used to exploit the fundamental coherence between fibre directions in neighbour voxels. Our method approaches the ℓ_0 minimisation through a reweighted ℓ_1 -minimisation scheme. The weights are here defined in such a way to promote correlated sparsity between neighbour signals.

Index Terms— structured sparsity, convex optimisation, diffusion MRI

1. INTRODUCTION

Diffusion Magnetic Resonance Imaging (dMRI) is an imaging modality which is sensitive to the Brownian motion of water molecules in living tissues. The anisotropic diffusion of water molecules can be measured to study the structure of spatial order in living organs in a non-invasive way. Since the axon membranes are the main responsible of the anisotropy of the water displacements in the white matter (WM), dMRI represents an effective technique to map the structural connectivity of the brain *in vivo*. By applying a pair of gradient pulses, MRI sequences become sensitive to motional processes, such as diffusion. These diffusion gradients are represented as 3D-vectors \vec{q} — oriented in the direction of the diffusion and with a magnitude proportional to the gradient strength— and they determine the 3D sampling space, commonly named after as q -space. Despite all its potential, the acquisition of diffusion MRI sequences is also known to be significantly time-consuming, restricting its use, for instance, in clinical practice. Thus, faster acquisitions, relying on as few q -sampling points as possible represents a major challenge in the field.

A great variety of approaches have been proposed for tackling the problem of intra-voxel fibre orientation estimation. In this paper, we focus on *spherical deconvolution* (SD)

methods [1–3] which have become very popular in the field of local reconstruction since they can recover the fibre configuration with a relatively small number of points. In [4], the reader can find the formal equations describing the relationship between the diffusion signal and the fibre orientation distribution function (FOD), a real-valued function on the unit sphere that indicates the orientation and the volume fraction of the fibre populations in a voxel. In a nutshell, the diffusion signal can be expressed as the convolution of a response function, or kernel, with the FOD as:

$$\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\eta}. \quad (1)$$

In (1), $\mathbf{x} \in \mathbb{R}_+^n$ stands for the FOD, $\mathbf{y} \in \mathbb{R}_+^m$ is the vector of measurements, Φ is the linear measurement operator (or dictionary) and $\boldsymbol{\eta}$ represents the acquisition noise. A set of the Diffusion Basis Functions [5] are a common choice to expand dictionary Φ . In that basis, each atom represents the diffusion profile corresponding to a single fibre rotated in a discrete set of n different directions equally distributed on the sphere. The components \mathbf{x}_i of our function of interest – the FOD – can be identified with the volume fraction of the fibre, oriented after the direction corresponding to the i -th atom of dictionary Φ . Therefore, by construction, \mathbf{x} is a non-negative vector with $\sum_i \mathbf{x}_i = 1$. At the current imaging resolution, at maximum 2–3 fibres per voxel can be distinguished [6]. Hence, the FOD can be considered a *sparse* function since the diffusion signal can be expressed as a linear combination of very few *atoms* of our dictionary. As a consequence and also in the light of the need of q -space subsampling, compressed sensing (CS) ideas [7, 8] have become popular in the field and sparsity priors are commonly used to regularise the ill-posed SD inverse problems.

Most of the state-of-the-art methods to solve SD problems use ℓ_1 minimisation [4, 5] as the common convex relaxation to promote sparsity on the solution. Yet, as pointed out in [9], the ℓ_1 minimisation is formally inconsistent with the fact that the volume fractions sum up to unity. A reweighted ℓ_1 minimisation scheme [10] was first proposed as an alternative for FOD recovery in [9]. It approaches ℓ_0 minimisation by a

sequence of convex weighted- ℓ_1 problems of the form:

$$\min_{\mathbf{x} \geq 0} \quad \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_{w,1} \leq k, \quad (2)$$

where $\|\cdot\|_{w,1}$ represents a weighted ℓ_1 norm defined as $\|\mathbf{x}\|_{w,1} = \sum_i w_i x_i$ and k acts as a bound on the expected number of fibre populations in a voxel. The iterative algorithm alternates between estimating the solution and redefining the weights essentially as the inverse of the values of the solution at the previous iteration so that, at convergence, the weighted ℓ_1 norm mimics the ℓ_0 norm [10]. Solving a sequence of problems (2), promotes sparsity in the solution while avoids working directly with the nonconvex ℓ_0 norm and overcomes the ℓ_1 inconsistency. However, formulation (2), as many other standard techniques to perform local reconstruction, focus on a voxel-by-voxel approach and thus, they are unable to exploit any spatial coherence of the data.

In this paper we review the method presented in [11], that incorporates structured sparsity ideas to constrain the reconstruction and simultaneously convey spatial information. Indeed, the assumption of a smooth evolution of fibre bundles in the brain can be directly translated into a coherence of the FOD in neighbouring voxels. In section 2, we show through numerical simulations how an adequate structured sparsity prior improves the quality on the solution and allows us to go to higher undersampling regimes. To conclude we describe the generic nature of our method, that can be leveraged in any application field for simultaneous recovery of correlated multiple sparse signals.

2. STRUCTURE THROUGH REWEIGHTING IN dMRI

2.1. Algorithm

In this section we consider the simultaneous recovery of multiple sparse signals $\mathbf{x}^{(v)}, v \in \{1, \dots, N\}$, corresponding to the FOD of N different voxels of the brain. By concatenating all vectors $\mathbf{x}^{(v)}$ column-wise, a matrix $\mathbf{X} \in \mathbb{R}_+^{n \times N}$ can be built, whose columns correspond to the individual signals. Inspired by formulation (2), we propose to resolve the FOD for all voxels simultaneously through a sequence of problems of the form:

$$\min_{\mathbf{X} \in \mathbb{R}_+^{n \times N}} \quad \|\Phi \mathbf{X} - \mathbf{Y}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{X}\|_{W,1} \leq K. \quad (3)$$

In (3), $\|\cdot\|_{W,1}$ stands for a weighted- ℓ_1 norm of a matrix defined as:

$$\|\mathbf{X}\|_{W,1} = \sum_{d,v} W_{dv} |\mathbf{X}_{dv}|, \quad (4)$$

where the weighting matrix W has the same dimension as \mathbf{X} and each of its entries, indexed with a pair (d, v) , acts as a weight for the corresponding entry of \mathbf{X} . The assumption

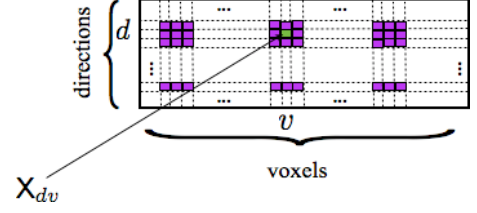


Fig. 1. Schematic representation of the neighbourhood (in purple) of element \mathbf{X}_{dv} (in green), i.e. the elements of \mathbf{X} involved in the computation of weight W_{dv} .

of smooth trajectories of the fibre tracts implies similar orientations of the maximal diffusion direction in neighbouring voxels. This premise can be directly translated to the FOD, since each of its coefficients is associated to one of the kernels of our dictionary, i.e. to a discrete direction on the sphere. As a consequence, there is a correlation in the support of the nonzero coefficients between the columns of \mathbf{X} , leading to a structure of its sparsity. A joint sparsity model [12, 13] would impose a common signal support, which does not apply in this context. Imposing low-rankness of \mathbf{X} through a nuclear norm minimisation [14] is not a better model as our signals could actually have maximally disjoint supports, leading to a high rank matrix. Our approach consists in meaningfully designing the weights of a reweighting scheme, not only to target ℓ_0 norm minimisation for each signal but also to impose that the signal associated with neighbour voxels should have neighbour orientation support.

To translate this idea into a mathematical formulation we introduce the concept of *neighbourhood*. For each atom of the dictionary d associated with a particular direction on the sphere, we define an *angular neighbourhood*, $\mathcal{N}(d)$, composed by its *closest atoms* in terms of angular distance. Analogously, for each voxel v we define its *spatial neighbourhood*, $\mathcal{N}(v)$, as the group of its 26 closest *voxels*, commonly referred to as the 26-adjacent neighbourhood [15]. We then define the neighbourhood of an element \mathbf{X}_{dv} as the entries of \mathbf{X} at the intersection of rows d and all its neighbour atoms, and columns v and all its neighbour voxels: $\mathcal{N}(dv) = \{(d', v'); d' \in (d \cup \mathcal{N}(d)), v' \in (v \cup \mathcal{N}(v))\}$, as it is schematically represented in Figure 1.

At each iteration t , every element of the weighting matrix W_{dv} is set as the inverse of an *average* of the absolute values that \mathbf{X} takes in the neighbourhood of \mathbf{X}_{dv} in the previous iteration:

$$W_{dv}^{(t+1)} = \left[\tau^{(t)} + \frac{1}{|\mathcal{N}(dv)|} \sum_{d',v' \in \mathcal{N}(dv)} |\mathbf{X}_{d'v'}^{(t)}| \right]^{-1}. \quad (5)$$

As a consequence, all entries corresponding to the neighbourhood of an element contribute to its weight. Therefore those orientations that are “supported” by the surrounding voxels are reinforced, since they will be given a small weight com-

pared to isolated ones that are not coherent with their environment. Fibre contributions are usually distributed over a small angular support since the true direction does in general not coincide exactly to one of the discrete points of the sphere identifying our orientation dictionary. Therefore, in (5), we average over voxels, but sum over directions as all values in neighbour directions are interpreted as corresponding to a single true local direction.

In (5), τ represents a stability parameter to avoid infinite values in the definition of the weights. In the absence of any spatial constraint, $W^{(0)}$ corresponds to the matrix of all 1s and thus, the weighted ℓ_1 norm is the standard $\ell_{1,1}$ norm of a matrix, $\|X\|_{W,1} = \|X\|_{1,1}$. At convergence, our definitions (4) and (5) implement a spatially coherent version of the matrix ℓ_0 norm, i.e. the sum of the ℓ_0 norms of its columns.

Each weighted- ℓ_1 problem of the form (3) is solved using Douglas-Rachford algorithm [16] in the context of proximal splitting theory [17]. The reader can find further details on the algorithm, as well as on the choice of the parameters (K, τ) in [11]. Hereafter, we refer to the sequence of problems (3) using the weights defined in (5) as $L2L0_{SS}$, in reference to the *structured sparsity* promoted. In the next subsection, we compare $L2L0_{SS}$ with $L2L0$, the voxel-by-voxel scheme (2), considered as a state-of-the-art SD method for sparse reconstruction of the FOD.

2.2. Numerical Simulations

In this subsection we compare in detail the performance for $L2L0_{SS}$ relative to $L2L0$ for a phantom data set¹ ($16 \times 16 \times 5$ volume) that comprises 5 different fibre bundles, resulting in voxels with different number of fibre populations (1 to 5) and crossing angles from 15° to 90° . The diffusion signal is contaminated with Rician noise [18] and the quality of the reconstructions is reported as a function of two different noise levels, i.e. $SNR = 10, 30$ and 5 different acquisition schemes corresponding to 30, 20, 15, 10 and 6 q -space samples per voxel.

To evaluate the quality of the reconstructions we focus on the performance of each method in both correctly assessing the number of fibre populations in each voxel and the angular accuracy in their orientation. The *success rate* (SR_L) corresponds to the proportion of voxels in which a reconstruction algorithm correctly estimates the number of fibre populations. The *angular accuracy* is measured through the *mean angular error* $\bar{\theta}$ (in degrees) over all true fibre directions. Note that index SR_L represents a mean value over all voxels of interest, whereas $\bar{\theta}$ is computed voxelwise and we study its statistical distribution to evaluate the general angular accuracy of each reconstruction. More details on the formal definition of these metrics can be found in [19].

¹Phantom data set used for the HARDI reconstruction Challenge 2012 <http://hardi.epfl.ch/static/events/2012/ISBI>

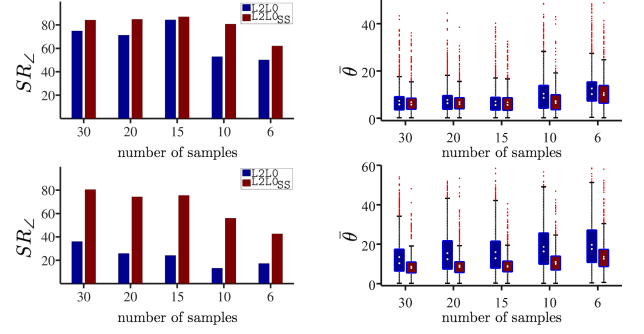


Fig. 2. Comparison of SR_L and $\bar{\theta}$ between $L2L0$ and $L2L0_{SS}$ approaches. Experiments are performed on a phantom dataset for a fixed $SNR = 30$ (top row) and $SNR = 10$ (bottom row). On the left, SR_L represents the *success rate*. On the right, the boxplot diagrams present the distribution of $\bar{\theta}$, with the edges of each box representing the 25th and 75th percentiles, the mean and median value appear as “square” and “circle” value and the outliers are plotted as red dots.

The performance of both methods as a function of the number of samples used for the reconstruction is reported in Figure 2. The plots demonstrate that $L2L0_{SS}$ outperforms $L2L0$ for all number of samples, in both noise conditions. At $SNR=30$, $L2L0_{SS}$ exhibits an accurate reconstruction ($SR_L \geq 85$ and $\text{mean}(\bar{\theta}) \leq 6.5^\circ$), robust to noise for different undersampling regimes, down to 15 samples. The plots show even a larger difference between the performance achieved by the two methods when we go to a more noisy situation ($SNR=10$). In particular, the overall $\bar{\theta}$ performances of both methods differ significantly, with an average enhancement of up to 5° in the mean $\bar{\theta}$ in different undersampling regimes.

$L2L0$ and $L2L0_{SS}$ have also been tested on a real dataset, corresponding to a whole brain volume, with $\approx 6 \cdot 10^6$ voxels and originally 256 q -samples per voxel. To assess the robustness of both algorithms to different undersampling rates, the dataset was undersampled and three additional datasets were created, consisting of only 30, 20 and 10 evenly spaced-distributed q -samples. The actual SNR was about 30. To evaluate the reconstructions from the subsampled real datasets, the metrics described formerly have been computed considering the fully-sampled dataset as the golden truth, as it is suggested in [20]. Results are in agreement with those obtained for numerical simulations on the phantom, confirming that $L2L0_{SS}$ actually outperforms $L2L0$. Figure 3 summarises the numerical results for one representative slice of the brain volume. The ability of correctly assessing the number fibre populations is illustrated through the number of over-/underestimated fibre populations per voxel (*false positive* and *negatives*, respectively). All detailed numerical results on real data can be found in [11].

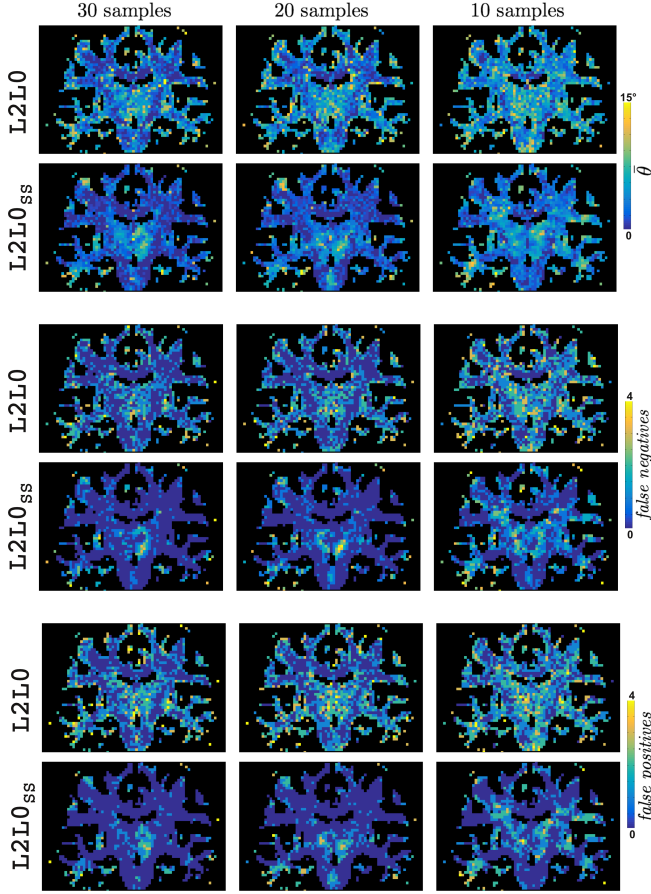


Fig. 3. Comparison in real data between $L2L0$ and $L2L0_{ss}$ reconstructions with 30, 20 and 10 samples. The quality of reconstruction is reported as angular accuracy (map of θ per voxel; top) and ability of correctly assessing the number of fibres (map of number of *false positives* and *false negatives* per voxel; middle and bottom, respectively).

3. DISCUSSION

So far in this paper we have presented a method to promote spatial regularisation through a structured sparsity prior and we have applied it for fibre orientation recovery from diffusion MRI data. We have shown through numerical simulations that this method outperforms a *simple sparse* regularisation on a voxel-by-voxel level when assessing the correct number of fibres and the angular precision of their orientation. As shown in section 2 for the problem in dMRI, promoting spatially structured sparsity guarantees a stronger robustness to noise and the ability to go to higher undersampling regimes.

Before concluding, we discuss the versatility of $L2L0_{ss}$ to be applied in a more generic framework for multiple correlated sparse signal recovery. In section 2, spatially-correlated vectors are concatenated to build matrix X . Formulation (3) can actually be generalised to recover multiple sparse signals correlated through a smoothness prior on the variation of the

signal support in the inter-signal dimension. Indeed, considering the concatenation of n_D *correlated* sparse signals into a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}$ and a linear operator Φ that models a measurement process on them, equation (3) can be rewritten as:

$$\min_{\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}} \|\Phi(\mathcal{X}) - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \|\mathcal{X}\|_{\mathcal{W},1} \leq K. \quad (6)$$

In (6), $\|\cdot\|_{\mathcal{W},1}$ represents a weighted- ℓ_1 norm of a generic tensor defined as:

$$\|\mathcal{X}\|_{\mathcal{W},1} = \sum_{i_1, \dots, i_D} \mathcal{W}_{i_1 \dots i_D} |\mathcal{X}_{i_1 \dots i_D}|. \quad (7)$$

The definition of the weights and neighbourhoods will enable the embedding of the smoothness prior in the signal support through the structure on the sparsity and must, of course, be adapted to the application. Our method stands in contrast to other joint sparsity models [12, 13] that assume a common support of the correlated signals. Social sparsity models [21, 22] also leverage the concept of neighbourhoods to promote sparsity. However, $L2L0_{ss}$ is essentially inspired from the reweighting scheme proposed by Candes et al. [10] to approach ℓ_0 minimisation through the convex minimisation of a weighted ℓ_1 norm. Our contribution lies in enforcing a *structured version* of the ℓ_0 norm of the solution at convergence and thus, the weights should continue to represent the inverse value of the associated entry at the previous iteration, to converge to an ℓ_0 norm.

In recent work [23] the authors propose an evolution of the recent Sparsity Averaging Reweighted Analysis approach (SARA) [24,25] for single signal recovery. In that context, the reweighting weights can be defined as aggregate values in entire blocks of a sparsifying dictionary, thus promoting structured sparsity. That reweighting scheme can be understood in the context of our approach, for neighborhoods defined as subsets of a sparsifying dictionary.

Future work in the same field of dMRI will consider the simultaneous recovery of the main directions and microstructure properties (axon diameter, for instance) of the fibre bundles in each voxel of the brain. That problem can be modelled as (6), where each of the dimensions of the tensor of interest would correspond to *voxel*, *orientation* and *axonal diameter*, respectively.

REFERENCES

- [1] J.D. Tournier, F. Calamante, D. Gadian, and A. Connelly, "Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution," *NeuroImage*, vol. 23, pp. 1176–1185, 2004.
- [2] D.C. Alexander, "Maximum entropy spherical deconvolution for diffusion MRI," *Information Processing in Medical Imaging (IPMI)*, pp. 76–87, 2005.

- [3] J.D. Tournier, F. Calamante, and A. Connelly, "Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution," *NeuroImage*, vol. 35, pp. 1459–1472, 2007.
- [4] B. Jian and B.C. Vermuri, "A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 1464–1471, 2007.
- [5] A. Ramirez-Manzanares, M. Rivera, B. Vemuri, P. Carney, and T. Mareci, "Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 1091–1102, 2007.
- [6] B. Jeurissen, A. Leemans, D. Jones, J. Tournier, and J. Sijbers, "Estimating the number of fiber orientations in diffusion MRI voxels : a constrained spherical deconvolution study," in *International Society for Magnetic Resonance in Medicine (ISMRM)*, p. 573, 2010.
- [7] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 54, pp. 1289–1306, 2006.
- [8] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 54, pp. 489–509, 2006.
- [9] A. Daducci, D. Van De Ville, J.P. Thiran, and Y. Wiaux, "Sparse regularization for fiber ODF reconstruction: from the suboptimality of ℓ_2 and ℓ_1 priors to ℓ_0 ," *Medical Image Analysis*, vol. 18, no. 6, pp. 820–33, 2014.
- [10] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.
- [11] A. Auría, A. Daducci, J.P. Thiran, and Y. Wiaux, "Structured sparsity for spatially coherent fibre orientation estimation in diffusion MRI," *NeuroImage*, pp. 245–255, 2015.
- [12] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *In Asilomar Conf. Signals, Sys., Comput*, 2005, pp. 1537–1541.
- [13] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE TRANS. SIGNAL PROCESS*, vol. 59, no. 9, pp. 4053–4085, 2011.
- [14] Ji. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.
- [15] J. Huang, R. Yagel, V. Filippov, and Y. Kurzion, "An accurate method for voxelizing polygon meshes," *Proceedings of the 1998 IEEE Symposium on Volume Visualization*, pp. 119–126, 1998.
- [16] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, Dec. 2007.
- [17] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point Algorithms for Inverse Problems in Science and Engineering (Bauschke H. H., Burachik R.S., Combettes P.L., Elser V., Luke D.R., Wolkowicz H., eds)*, pp. 185–212, 2011.
- [18] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 910–914, 1995.
- [19] A. Daducci, E.J. Canales-Rodriguez, M. Descoteaux, E. Garyfallidis, Y. Gur, Y. Lin, M. Mani, S. Merlet, M. Paquette, A. Ramirez-Manzanares, M. Reisert, P.R. Rodrigues, F. Sepehrband, E. Caruyer, J. Choupan, R. Deriche, M. Jacob, G. Menegaz, V. Prckovska, M. Rivera, Y. Wiaux, and J.P. Thiran, "Quantitative Comparison of Reconstruction Methods for Intra-Voxel Fiber Recovery From Diffusion MRI," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 384–399, 2014.
- [20] F.C. Yeh and W.Y. Tseng, "Sparse solution of fiber orientation distribution function by diffusion decomposition," *PLoS One*, vol. 8, 2013.
- [21] M. Kowalski and B. Torrèsani, "Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients.," *Signal, Image and Video Processing*, vol. 3, no. 3, pp. 251–264, 2009.
- [22] M. Kowalski, K. Siedenburg, and M. Dörfler, "Social Sparsity! Neighborhood Systems Enrich Structured Shrinkage Operators.," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2498–2511, 2013.
- [23] Rizwan Ahmad and Philip Schniter, "Iteratively reweighted ℓ_1 approaches to sparse composite regularization," .
- [24] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (sara): a novel algorithm for radio-interferometric imaging," *MNRAS*, vol. 426, no. 2, pp. 1223–1234, 2012.
- [25] R.E. Carrillo, J.D. McEwen, D. Van De Ville, J.-P. Thiran, and Y. Wiaux, "Sparsity averaging for compressive imaging," *IEEE Signal Processing Letters*, vol. 20, no. 6, pp. 591–594, 2013.